KERATIN 8 AND 18 MUTATIONS ARE RISK FACTORS FOR DEVELOPING LIVER DISEASE OF MULTIPLE ETIOLOGIES

INTRODUCTION

[01] Keratin mutations are associated with several skin, oral, esophageal, ocular and cryptogenic liver diseases that reflect tissue-specific expression of the particular keratin involved. The resulting cellular and tissue defects are manifestations of the clearly defined function of keratins that provides cells with the ability to cope with mechanical stresses. This keratin cytoprotective effect is evident in the blistering phenotype of several human keratin skin diseases such as epidermolysis bullosa simplex, and the phenotypes of animal models that lack or express a mutant keratin. Also, emerging evidence suggests that keratins protect cells from nonmechanical forms of injury via several mechanisms that may include: keratin regulation of cell signaling cascades, regulation of apoptosis, regulation of the availability of other cellular proteins, and protein targeting to subcellular compartments.

[02] The function of keratins in protecting cells from mechanical stress is related to their unique properties and abundance as one of three major cytoskeletal protein families, which include intermediate filaments (IF), microfilaments and microtubules. Keratins (K) are members of the IF protein family, and are specifically expressed in epithelial cells and their appendages. They consist of >20 members (K1-K20), and are further classified into type I (K9-K20) and type II (K1-K8) keratins which form obligate, noncovalent heteropolymers. Keratins serve as important cell-type-specific markers. For example, unique keratin complements distinguish different epithelial cell types and thereby reflect epithelial subtype-specific diseases that result from keratin-specific mutations. As such, keratinocytes express K5/K14 basally and K1/K10 suprabasally, and hepatocytes express K8/K18. K8/K18 are also found in other glandular cells including enterocytes, with variable complements of K19/K20/K7 depending on the cell type.

[03] Most keratin diseases are autosomal-dominant with near complete penetrance. Exceptions appear to be K18 and K8 mutations in patients with cryptogenic cirrhosis. To date, 6 patients have been described with K8 (5 patients) or K18 (1 patient) mutations, from a group of 55 patients with cryptogenic cirrhosis. Most patients with cryptogenic cirrhosis, including those with K8/K18 mutations, do not have a well-defined liver disease family history. Absence of a clear family history suggests that K8/K18 mutations predispose to, rather than cause, liver disease. The presence and frequency of keratin mutations in noncryptogenic liver disease is heretofore unknown.

Summary of the Invention

[04] Keratin 8 and 18 (K8/K18) mutations are shown to be associated with a predisposition to liver disease, particularly noncryptogenic liver disease. Unique K8/K18 mutations are shown in patients with diseases including biliary atresia, acute fulminant hepatitis, viral hepatitis B or C, alcoholic liver disease, primary biliary cirrhosis, autoimmune hepatitis, and the like. Livers with keratin mutations had increased incidence of cytoplasmic filamentous deposits. Therefore, K8/K18 are susceptibility genes for developing cryptogenic and noncryptogenic forms of liver disease. Alleles are associated with disease susceptibility, and their detection is used in the diagnosis of a predisposition to these conditions.

[05] The invention also provides methods for the identification of compounds that modulate the expression of genes or the activity or the cellular organization or the organizational stability of keratin gene products involved in hepatobiliary disease, as well as methods for the treatment of hepatobiliary disease, which may involve the administration of such compounds to individuals exhibiting liver disease symptoms or tendencies. Similar compounds may also be used to treat other keratin or other intermediate filament related diseases.

Brief Description of the Drawings

[06] Fig. 1: Protein domains analyzed for K8/K18 mutations and an example of the identification of a K8 mutation. (A) A central rod domain, consisting of α-helical subdomains, is flanked by non-α-helical head/tail domains. The head/tail domains are further subdivided into E, V and H regions. The subdomains of the rod are connected by nonhelical linker (L1, L1-2, L2) regions. The amino acid (aa) regions in black bars represent 5 domains that were examined for K8/K18 mutations. The remaining regions in gray bars contain 10 exonic K8/K18 domains that have been analyzed for mutations. (B) PCR products, from a control patient (with K8 WT) and a patient with K8 R340H, were analyzed by denaturing HPLC using a WAVE® System. The control (K8 WT) is characterized by one major peak, while the K8 R340H shows a different chromatogram due to resolution of the homoduplexes from the heteroduplexes, thereby suggesting the presence of a K8 mutation. Electropherograms from DNA sequencing confirm the presence of a K8 R340H heterozygous missense mutation (CGT→CAT).

[07] Fig. 2: Protein expression of mutant K8 and K18 in explanted livers. (A): K8/K18 immunoprecipitates were obtained from 1% Empigen-solubilized normal liver or livers with keratin mutation. The immunoprecipitates were separated by isoelectic focusing followed by SDS-PAGE, then immunoblotting with anti-K8/K18 antibodies. Note that K8 and K18 in normal liver consist of two or three isoforms depending on their phosphorylation levels (a, b). In contrast, some of the mutant keratins contain four (K8) or five (K18) isoforms due to coexpression of the wild-type and mutant keratin with subsequent generation of altered

charged species that have a slightly different mutation-induced isoelectric focusing point (d, f, g, h). (B): K8/K18 immunoprecipitates were prepared from normal liver or liver with the K18 T102A mutation, then analyzed by SDS-PAGE. The K18 bands were cut out, digested with trypsin, then analyzed with a MALDI-TOF mass spectrometer. Note that a peak position at 818.3 was detected only in the liver specimen with the K18 T102A mutation but not in normal liver. The mass difference of 30, between the wild-type and T102A K18 tryptic peptides (848.3 versus 818.3), corresponds to the HO-C-H species (two hydrogen, one oxygen and one carbon atoms with a mass of 30 daltons) that are present in threonine (the wild-type residue) but not in alanine (the mutant residue). ...

[08] Fig. 3: K8 R340H mutation-proximal comparison of type II keratin sequences and confirmation of K8 R340H mutant protein expression in explanted livers. (A) Single letter abbreviations are used to represent amino acids. Bold dots represent amino acids that are identical to the K8 sequence. The shaded area highlights the conserved R340 of K8 and shows the histidine mutation we identified. Note that the K8 R340-containing motif (AEQ*R*GE) is highly conserved in type II keratins. It is also conserved across species, being found in mouse and frog K8. (B) BHK cells were transiently cotransfected with K8/K18 WT or K8 R340H/K18 WT. K8/K18 immunoprecipitates were obtained from 1% NP40-solubilized cell lysates. The immunoprecipitates were analyzed by SDS-PAGE, followed by immunoblotting with anti-K8 R340 or anti-K8 H340 epitope-specific antibodies that preferentially recognized K8 WT or K8 R340H mutant, respectively. (C) K8/K18 immunoprecipitates were obtained from 1% NP40-solubilized normal liver or livers with the K8 R340H mutation. Samples were analyzed as described in panel B. Note that anti-K8 R340 (WT) recognizes K8 in the control patient (with K8 WT) and the patient with K8 R340H (lanes 1-6), whereas anti-K8 H340 (mutant) recognizes only patient livers with the K8 R340H mutation (lanes 2-6) but not the normal liver (lane 1). This indicates that the patients with K8 R340H mutation are heterozygous with regard to the keratin mutation. Arrowheads correspond to degraded K8.

[09] Fig. 4: Keratin filament organization in human liver explants, and histologic findings of livers harboring the keratin mutations. (A): Human livers were sectioned, fixed in acetone and double-stained with rabbit anti-K8/18 or mouse anti-vimentin antibodies. Inset in panel i shows control double staining using fluorochrome-conjugated goat anti-rabbit and goat anti-mouse antibodies without adding any primary antibodies. All images were obtained using the same magnification. Bar in panel a = 20 μm. (B): Hematoxylin and eosin staining of explanted liver from two patients with acute fulminant hepatitis. Panel "a" is from a patient without a keratin mutation while panel "b" is from a patient with the K18 T102A mutation. The region outlined by a box in "b" is magnified in panel "c" to illustrate the cytoplasmic filamentous deposits noted primarily in livers of patients with keratin mutations.

[10] Fig. 5: Distribution of K8 and K18 mutations within the keratin protein backbone. The H1 and V1 subdomains of the K8 head domain, and the amino acid positions of the K8 and K18 domains/subdomains are indicated. For the K8 head/tail and K18 head/rod domains, each arrowhead represents an independently identified mutation at the indicated residue. The 30 K8 R340H mutations are highlighted with a single large arrowhead. Note that the most common K8/K18 mutation to date is K8 R340H (30 of 58, ~52%), with the next most common mutations being K8 G61C (6 of 58) and K8 Y53H (5 of 58). Three patients had a double mutant (K8 R340H and K18 R260Q, K8 R340H and K18 T102A, K18 I149V and K18 T294M). Shaded areas correspond to "hot spots" of epidermal keratin mutations. Note the non-overlap of the location of K8/K18 mutations versus the major location of epidermal keratin mutations.

[11] Fig. 6: Schematic model of the effect of keratin mutations on filament organization and subsequent liver injury. Hepatocytes contain an extended array of filamentous keratin and a small soluble keratin pool. A variety of physiologic and nonphysiologic stimuli may result in reversible keratin filament reorganization. However, in the presence of keratin mutations, the keratin filaments may be unstable (as for the K8 Y53H mutation upon heat stress or exposure to okadaic acid), or may not be able to organize normally (as for the K8 G61C mutation upon oxidative stress; e.g. in the presence of H_2O_2). Filament instability and abnormal organization may then predispose to liver injury via a variety of mechanisms that include the role of keratins in apoptosis, cell signaling or protein targeting to subcellular compartment.

[12] The above six figures demonstrate: (i) the presence of the keratin mutations at the protein level (Fig. 2 and 3), (ii) although immunofluorescence staining is not significantly altered in the livers with keratin mutation as compared to those without keratin mutation (Fig. 4A) there appears to be a unique histologic feature of cytoplasmic filamentous deposits that are observed primarily in livers of patients with keratin mutations (Fig. 4B), (iii) distribution of K8 and K18 mutations in the keratin protein backbone (Fig. 5) and (iv) schematic model of the effect of keratin mutations on filament organization and subsequent liver injury (Fig. 6).

[13]

DETAILED DESCRIPTION OF THE EMBODIMENTS

[14] Methods and compositions are provided for the diagnosis of a predisposition to liver or biliary tract disease, including, without limitation, viral hepatitis, autoimmune liver disease, biliary atresia, alcoholic cirrhosis and other acute or chronic toxic liver injury, cryptogenic cirrhosis, acute fulminant hepatitis, primary sclerosing cholangitis, primary biliary cirrhosis, diseases that are linked with cryptogenic cirrhosis, such as nonalcoholic steatohepatitis, and the like. The invention is based, in part, on the evaluation of the K8/K18 keratin genotype, for which alleles predisposing to disease are herein identified. This permits the definition of disease pathways and the identification of a target in the pathway that is useful diagnostically, in drug screening, and therapeutically.

[15] In one aspect of the present invention, methods are provided for determining a predisposition to liver disease in an individual. The methods comprise an analysis of genomic DNA in an individual for an allele of keratin K8 or K18 that confers an increased susceptibility to liver disease. Individuals are screened by analyzing their genomic K8 and/or K18 gene sequence, e.g. in blood; tissue/cell specimen; etc. for the presence of a predisposing allele, as compared to a normal sequence. Screening for the presence of the mutation can also be done using antibodies that specifically identify the keratin mutation, or screening of genomic material in serum.

[16] In addition to the provided sequence polymorphisms, the effect of a candidate polymorphism in a K8 or K18 sequence can be determined for association with a predisposition to liver disease. The candidate polymorphism may be analyzed, for example, for segregation of the sequence polymorphism with the disease phenotype. A predisposing mutation will segregate with incidence of the disease. Alternatively, biochemical studies may be performed to determine whether a candidate polymorphism affects the quantity or quality (in terms of its distribution, interaction with binding partners, etc), or function of the protein. As used herein, the term polymorphism is generally used to refer to polypeptide or polynucleotide sequences in which two or variants are present in a population. The term mutation is used herein to refer to a polymorphism that has an association with disease.

[17] Intermediate filaments (IFs) are a structurally related family of cellular proteins that are intimately involved with the cytoskeleton. The common structural motif shared by all IFs is a central alpha-helical 'rod domain' flanked by variable N- and C-terminal domains. The rod domain, the canonical feature of IFs, has been highly conserved during evolution. The variable terminals, however, have allowed the known IFs to be classified into 6 distinct types by virtue of their differing amino acid sequences. Keratins compose types I and II IFs. Type I and type II keratins are usually expressed as preferential pairs, in equal proportions in cells, of type I and type II keratins.

[18] Human keratin 18 is a type I IF protein subunit, whose expression is restricted in adults to a variety of so-called "simple-type" epithelial tissues. KRT18 is highly divergent among the type I keratins with N-terminal and C-terminal domains that are quite different from those of epidermal keratins. The KRT18 gene is 3,791 bp long and the keratin 18 protein is coded for by 7 exons. The genetic sequence of K18 may be found in Genbank, accession no. NM_000224. For convenience, the polynucleotide sequence of K18 is provided as SEQ ID NO:1, and the amino acid sequence as SEQ ID NO:2. The exon structure of KRT18 has been conserved compared to that of other keratin genes, with the exception of a single 3-prime terminal exon that codes for the tail domain of the protein that is represented by 2 exons in epidermal keratins. Keratin 8 is a type II keratin, which is co-expressed with K18. The genetic sequence of K8 may be found in Genbank, accession no. NM_002273. For convenience, the

polynucleotide sequence of K8 is provided as SEQ ID NO:3, and the amino acid sequence as SEQ ID NO:4.

[19] The term "identical" in the context of two nucleic acid or polypeptide sequences refers to the residues in the two sequences, which are the same when aligned for maximum correspondence. When percentage of sequence identity is used in reference to proteins or peptides it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g. charge or hydrophobicity) and therefore may not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., according to known algorithm. See, e.g., Meyers and Miller, Computer Applic. Biol. Sci., 4: 11-17 (1988); Smith and Waterman (1981) Adv. Appl. Math. 2: 482; Needleman and Wunsch (1970) J. Mol. Biol. 48: 443; Pearson and Lipman (1988) Proc. Natl. Acad. Sci. USA 85: 2444; Higgins and Sharp (1988) Gene, 73: 237-244 and Higgins and Sharp (1989) CABIOS 5: 151-153; Corpet, et al. (1988) Nucleic Acids Research 16, 10881-90; Huang, et al. (1992) Computer Applications in the Biosciences 8, 155-65, and Pearson, et al. (1994) Methods in Molecular Biology 24, 307-31. Alignment is also often performed by inspection and manual alignment. "Conservatively modified variations" of a particular nucleic acid sequence refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance, the codons CGU, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of "conservatively modified variations." Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a polypeptide is implicit in each described sequence. Furthermore, one of skill will recognize that individual

substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative amino acid substitutions providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Serine (S), Threonine (T); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W). The term "complementary" means that one nucleic acid molecule has the sequence of the binding partner of another nucleic acid molecule. Thus, the sequence 5'-ATGC-3' is complementary to the sequence 5'-GCAT-3'.

[20] An amino acid sequence or a nucleotide sequence is "substantially identical" or "substantially similar" to a reference sequence if the amino acid sequence or nucleotide sequence has at least 80% sequence identity with the reference sequence over a given comparison window. Thus, substantially similar sequences include those having, for example, at least 85% sequence identity, at least 90% sequence identity, at least 95% sequence identity or at least 99% sequence identity. Two sequences that are identical to each other are, of course, also substantially identical.

[21] It will be understood by those of skill in the art that by referring to the polynucleotides and polypeptides according to the published sequence, the presence of other minor polymorphisms, e.g. silent and conservative sequence changes is also encompassed. Generally, the mutations described herein will be present in the context of a sequence that is substantially identical to a native K8 or K18 sequence.

[22] Mutations in K8 that are associated with a predisposition to liver disease (all the amino acid numbers represent amino acids of the processed protein) include G52V (GGC→GTC); K8 Y53H (TAT→CAT); K8 G61C (GGC→TGC); K8 R340H (CGT→CAT); K8 G433S (GGC→AGC); K8 R453C (CGC→TGC); K8 1-465(I)RDT(468) (frameshift). The K8 frameshift mutation at Ile-465 generates a truncated 468 (instead of 482) amino acid protein that contains amino acids 1-465, and three additional new amino acids (RDT) after the frameshift mutation.

[23] Mutations in K18 that are involved with a predisposition to liver disease include K18 Δ64-71(TGIAGGLA) (deletion); K18 T102A (ACC→GCC); K18 H127L (CAT→CTT); K18 I149V (ATC→GTC); K18 R260Q (CGG→CAG); K18 E275G (GAG→GGG); K18 Q284R (CAG→CGG); K18 T294M (ACG→ATG); K18 T296I (ACA→ATA); K18 G339R (GGG→AGG). The K18 N-terminal domain deletion (Δ64-71) generated a 421 (instead of 429) amino acid protein. A list of mutations is also provided in Table 3.

[24] Keratin K8 or K18 mutations that result in an amino acid substitution or deletion at any one of the mutated positions designated above may be generally defined to include K8 G52X; Y53X; G61X; R340X; G433X; R453X and K18 T102X; H127X; I149X; R260X; E275X; Q284X; T294X; T296X; G339X, where X is any amino acid other than the naturally occurring amino acid as set forth in the published sequence of K8 or K18, and X may also refer to a deleted amino acid or amino acids. Mutations may also be specified in terms of the DNA sequence, and could include any deletions (in addition to those described above), or mutations in the promoter or other regulatory regions that may affect RNA levels or stability under basal conditions or in response to any type of mechanical or nonmechanical stress that the liver may be exposed to due to internal or external factors.

[25] DNA encoding a K8/K18 protein may be cDNA, genomic DNA, a fragment thereof, primer, *etc.* that encompasses one or more of the above identified polymorphisms (except for native sequences, which may find use as controls). As known in the art, cDNA sequences have the arrangement of exons found in processed mRNA, forming a continuous open reading frame, while genomic sequences may have introns interrupting the open reading frame. The term K8/K18 gene shall be intended to mean the open reading frames encoding such specific polypeptides, as well as adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 1 kb beyond the coding region, in either direction. Fragments may be obtained of the genetic sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.* For the most part, polynucleotide fragments will be at least about 20 nt in length, usually at least about 25 nt, at least about 50 nt, or more.

[26] Pairs of primer sequences used for amplification of a target sequence may flank the target mutation without encompassing it. For example, in order to amplify a mutation at amino acid residue 340 of K8, primers are designed to amplify the polynucleotide sequence at polynucleotide residue 1081. Such primers may be chosen from any of the 5' and 3' flanking regions, e.g. one might select residues SEQ ID NO:3, 1000-1020 for one primer; and for the complementary strand, SEQ ID NO:3, 1100-1120. Such selection of PCR primers employs methods well-known in the art. It will be understood that many combinations of primers may be used, as long as the amplification product encompasses the mutated sequence of interest. While the exact composition of the primer sequences is not critical to the invention, for most applications the primers will hybridize to the subject sequence under conditions suitable for amplification, as known in the art. It is preferable to chose a pair of primers that will generate an amplification product of at least about 25 nt, at least 50 nt, at least about 100 nt or more. Algorithms for the selection of primer sequences are generally known, and are available in commercial software packages. Amplification primers hybridize to complementary strands of DNA, and will prime towards each other.

[27] Primers for direct detection of the mutation will usually encompass the mutated sequence of interest, and will comprise flanking sequences, *e.g.* about 5, about 10, about 15 or more nucleotides 5' and 3' to the mutated residue. For example, a primer designed to bind to the K8 R340H sequence may extend from SEQ ID NO:3 position 1071 – 1091, so long as the primer is sufficiently long as to selectively bind to the sequence of interest.

[28] The subject K8/K18 genes are isolated and obtained in substantial purity, generally as other than an intact mammalian chromosome. Usually, the DNA will be obtained substantially free of other nucleic acid sequences that do not include a K8/K18 sequence or fragment thereof, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", i.e. flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome. The subject nucleic acids may be used to identify expression of the gene in a biological specimen.

[29] A number of methods may be used for determining the presence of a predisposing mutation in an individual. Genomic DNA may be isolated from the individual or individuals that are to be tested. DNA can be isolated from any nucleated cellular source such as blood, hair shafts, saliva, mucous, biopsy, feces, *etc.* mRNA may also be reverse transcribed from a tissue sample, depending on the detection method. Methods using PCR amplification can be performed on the DNA from a single cell, although it is convenient to use at least about 10^5 cells. Where large amounts of DNA are available, the genomic DNA is used directly. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis, or amplified by conventional techniques. Of particular interest is the use of the polymerase chain reaction (PCR) to amplify the DNA that lies between two specific primers. The use of the polymerase chain reaction is described in Saiki *et al.* (1985) Science 239:487, and a review of current techniques may be found in McPherson *et al.* (2000) PCR (Basics: From Background to Bench) Springer Verlag; ISBN: 0387916008. A detectable label may be included in the amplification reaction. Suitable labels include fluorochromes, *e.g.* fluorescein isothiocyanate (FITC); rhodamine; Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'- dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), radioactive labels, e.g. ^{32}P, ^{35}S, ^{3}H; *etc.* The label may be a two stage system, where the amplified DNA is conjugated to biotin, haptens, *etc.* having a high affinity binding partner, *e.g.* avidin, specific antibodies, *etc.*, where the binding partner is conjugated to a detectable label. The label may be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

[30] Primer pairs are selected from the K8/K18 genomic sequence using conventional criteria for selection. The primers in a pair will hybridize to opposite strands, and will collectively flank the region of interest. The primers will hybridize to the complementary sequence under stringent conditions, and will generally be at least about 16 nt in length, and may be 20, 25 or 30 nucleotides in length. The primers will be selected to amplify the specific region of the K8/K18 gene suspected of containing the predisposing mutation, usually by selecting primers that are 5' and 3' to the polynucleotides encoding the mutation. Multiplex amplification may be performed in which several sets of primers are combined in the same reaction tube, in order to analyze multiple exons/introns/promoter and other regulatory regions simultaneously. Each primer may be conjugated to a different label.

[31] After amplification, the products may be size fractionated and evaluated for sequence polymorphisms. Fractionation may be performed by gel electrophoresis, particularly denaturing acrylamide or agarose gels. A convenient system uses denaturing polyacrylamide gels in combination with an automated DNA sequencer, see Hunkapillar et al. (1991) Science 254:59-74. The automated sequencer is particularly useful with multiplex amplification or pooled products of separate PCR reactions. Capillary electrophoresis may also be used for fractionation. A review of capillary electrophoresis may be found in Landers, et al. (1993) BioTechniques 14:98-111. Hybridization with the variant sequence may also be used to determine its presence, by Southern blots, dot blots, *etc.* Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), and heteroduplex analysis in gel matrices is used to detect conformational changes created by DNA sequence variation as alterations in electrophoretic mobility or via HPLC type analysis (eg WAVE® System). The hybridization pattern of a control and variant sequence to an array of oligonucleotide probes immobilized on a microarray, may also be used as a means of detecting the presence of variant sequences.

[32] The presence of a predisposing mutation is indicative that an individual is at increased risk of developing liver disease. The diagnosis of a disease predisposition allows the affected individual to seek early treatment, dietary measures, to avoid activities that increase risk for liver disease, and the like.

[33] In addition to genetic tests, the presence of the mutated polypeptide may be detected, by determination of the presence of polypeptide comprising the mutation using analytic methods such as mass spectrometry or immune-related methods such as mutant-specific antibodies, or by detecting the presence of cytoplasmic filamentous deposits.

[34] In a typical assay, a liver sample is assayed for the presence of K8 and/or K18 specific sequences by combining the sample with a K8 and/or K18 specific binding member, and detecting directly or indirectly the presence of the complex formed between the two members. The term "specific binding member" as used herein refers to a member of a specific binding

pair, i.e. two molecules where one of the molecules through chemical or physical means specifically binds to the other molecule. In this particular case one of the molecules is K8 and/or K18, where K8 and/or K18 is any protein substantially similar to the amino acid sequence of the human polypeptide sequences of this family, as described above, or a epitope containing fragment thereof, and further comprising a predisposing mutation. The complementary members of a specific binding pair are sometimes referred to as a ligand and receptor. Testing can also be done using any source of genomic material from a given individual or any tissue that also expresses K8 and K18 (eg gastric, small or large intestinal biopsy).

[35] In the present specification and claims, the term "polypeptide fragments", or variants thereof, denotes both short peptides with a length of at least two amino acid residues and at most 10 amino acid residues, oligopeptides with a length of at least 11 amino acid residues, 20 amino acid residues, 50 amino acid residues, and up to about 100 amino acid residues; and longer peptides of greater than 100 amino acid residues up to the complete length of the native polypeptide.

[36] Polypeptides detected by the present methods include naturally occurring alpha and beta subunits, as well as variants that are encoded by DNA sequences that are substantially homologous to one or more of the DNA sequences specifically recited herein, for example variants having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% sequence identity. The specific binding pairs may include analogs, derivatives and fragments of the original specific binding member. For example, an antibody directed to a protein antigen may also recognize peptide fragments, chemically synthesized peptidomimetics, labeled protein, derivatized protein, etc. so long as an epitope is present.

[37] Immunological specific binding pairs include antigens and antigen specific antibodies or T cell antigen receptors. Recombinant DNA methods or peptide synthesis may be used to produce chimeric, truncated, or single chain analogs of either member of the binding pair, where chimeric proteins may provide mixture(s) or fragment(s) thereof, or a mixture of an antibody and other specific binding members. Antibodies and T cell receptors may be monoclonal or polyclonal, and may be produced by transgenic animals, immunized animals, immortalized human or animal B-cells, cells transfected with DNA vectors encoding the antibody or T cell receptor, etc. The details of the preparation of antibodies and their suitability for use as specific binding members are well-known to those skilled in the art.

[38] Alternatively, monoclonal or polyclonal antibodies are raised to K8 and/or K18 polypeptides comprising a predisposing mutation. The antibodies may be produced in accordance with conventional ways, immunization of a mammalian host, e.g. mouse, rat, guinea pig, cat, rabbit, dog, etc., fusion of resulting splenocytes with a fusion partner for immortalization and screening for antibodies having the desired affinity to provide monoclonal

antibodies having a particular specificity. These antibodies can be used for affinity chromatography, ELISA, RIA, and the like. The antibodies may be labeled with radioisotopes, enzymes, fluorescers, chemiluminescers, or other label, which will allow for detection of complex formation between the labeled antibody and its complementary epitope. Generally the amount of bound K8 and/or K18 detected will be compared to negative control samples from normal tissue or cells.

[39] Screening assays identify compounds that modulate the expression or structure of K8/K18, including compounds that stabilize these and other keratins or other intermediate filament proteins. Such compounds may include, but are not limited to peptides, antibodies, or small organic or inorganic compounds. Methods for the identification of such compounds are described below.

[40] Cell- and animal-based systems can act as models for liver disease and are useful in such drug screening. The animal- and cell-based models may be used to identify drugs, pharmaceuticals, therapies and interventions that are effective in treating liver disease. In addition, such animal models may be used to determine the LD_{50} and the ED_{50} in animal subjects, and such data can be used to determine the *in vivo* efficacy of potential liver disease treatments. Animal-based model systems of liver disease may include, but are not limited to, non-recombinant and engineered transgenic animals. Animal models exhibiting liver disease symptoms may be engineered by utilizing, for example, K8 and/or K18 gene sequences in conjunction with techniques for producing transgenic animals that are well known to those of skill in the art. For example, target gene sequences may be introduced into, and knocked out or overexpressed in the genome of the animal of interest. Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees may be used to generate liver disease animal models.

[41] Any technique known in the art may be used to introduce a target gene transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to pronuclear microinjection (Hoppe, P. C. and Wagner, T. E., 1989, U.S. Pat. No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten et al., 1985, Proc. Natl. Acad. Sci., USA 82:6148-6152); gene targeting in embryonic stem cells (Thompson et al., 1989, Cell 56:313-321); electroporation of embryos (Lo, 1983, Mol Cell. Biol. 3:1803-1814); and sperm-mediated gene transfer (Lavitrano et al., 1989, Cell 57:717-723); *etc.*

[42] Specific cell types within the animals may be analyzed and assayed for cellular phenotypes characteristic of liver disease. Further, such cellular phenotypes may include a particular cell type's fingerprint pattern of expression as compared to known fingerprint expression profiles of the particular cell type in animals exhibiting liver disease symptoms.

[43] Cells that contain and express K8/K18 can be utilized to identify compounds that exhibit pharmacologic activity of interest, in the prevention of liver disease. Cells of a cell type known to be involved in liver disease may be transfected with sequences capable of increasing or decreasing the amount of K8 and/or K18 gene expression within the cell. For example, K8/K18 gene sequences may be introduced into, and overexpressed in, the genome of the cell of interest, or, if endogenous target gene sequences are present, they may be either overexpressed or, alternatively disrupted in order to underexpress or inactivate target gene expression.

[44] Transfection of target gene sequence nucleic acid may be accomplished by utilizing standard techniques. Transfected cells can be evaluated for the presence of the recombinant K8/K18 gene sequences, for expression and accumulation of K8/K18 gene mRNA, and for the presence of recombinant K8/K18 protein. Where a decrease in K8/K18 gene expression is desired, standard techniques may be used to demonstrate whether a decrease in expression is achieved.

[45] *In vitro* systems may be designed to identify compounds capable of preventing or treating liver disease. Such compounds may include, but are not limited to, peptides made of D-and/or L-configuration amino acids, phosphopeptides, antibodies, and synthetic or natural small organic or inorganic molecules. For example, assays may be used to identify compounds that improve liver function involves preparing a reaction mixture of K8/K18 and a test compound under conditions and for a time sufficient to allow the two components to interact, and detecting the resulting change in the polypeptide structure. Alternatively, a simple binding assay can be used as an initial screening method. These assays can be conducted in a variety of ways. For example, one method to conduct such an assay would involve anchoring K8/K18 protein or a test substance onto a solid phase and detecting complexes anchored on the solid phase at the end of the reaction. In another embodiment of such a method, the assay tests the presence of products modulated by K8/K18.

[46] In a binding assay, the reaction can be performed on a solid phase or in liquid phase. In a solid phase assay, the nonimmobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously nonimmobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously nonimmobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the previously nonimmobilized component (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody).

[47] Alternatively, a binding reaction can be conducted in a liquid phase, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for target gene product or the test compound to anchor any complexes formed in solution, and a labeled antibody specific for the other component of the possible complex to detect anchored complexes.

[48] Cell-based systems such as those described above may be used to identify compounds that act to ameliorate liver disease symptoms. For example, such cell systems may be exposed to a test compound at a sufficient concentration and for a time sufficient to elicit such an amelioration of liver disease symptoms in the exposed cells. After exposure, the cells are examined to determine whether one or more of the liver disease cellular phenotypes has been altered to resemble a more normal or more wild type, non-liver disease phenotype. Liver disease "symptoms" in cells include surrogate markers such as changes in keratin filament organization, keratin properties (such as phosphorylation or solubility), or interaction with a binding partner.

[49] In addition, animal-based disease systems, such as those described, above may be used to identify compounds capable of ameliorating disease symptoms. Such animal models may be used as test substrates for the identification of drugs, pharmaceuticals, therapies, and interventions, which may be effective in treating or preventing disease. For example, animal models may be exposed to a compound, suspected of exhibiting an ability to ameliorate liver disease symptoms, at a sufficient concentration and for a time sufficient to elicit such an amelioration of disease symptoms in the exposed animals. The response of the animals to the exposure may be monitored by assessing the reversal of disorders associated with disease.

[50] With regard to intervention, any treatments that reverse any aspect of liver disease symptoms may be considered as candidates for human disease therapeutic intervention. Dosages of test agents may be determined by deriving dose-response curves.

[51] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals; e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50} /ED_{50}. Compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[52] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies

preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[53] Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvates may be formulated for administration by oral, buccal, parenteral or rectal administration.

[54] In addition to use for the treatment of liver disease, compounds that stabilize keratins may also find use in the treatment of other keratin-associated diseases. All IF proteins share a prototype structure of a coiled-coil α -helix "rod" domain, interrupted by linkers, that is flanked by N-terminal "head" and C-terminal "tail" domains. The simplest soluble unit of IF proteins is a tetramer that consists of two anti-parallel dimers with each dimer, in the case of keratins, consisting of one type I and one type II keratin. The mature 10-12 nm fiber is believed to consist of ~32 monomers across, organized into four intertwined protofibrils. There are many IF-related diseases that manifest as tissue, cell or nuclear fragility in skin, cornea, muscle, nerves, *etc.* Compounds that stabilize intermediate filament proteins (keratin and non keratin) may find broad use in the treatment of associated disorders.

[55] In one aspect, diseases associated with keratins are treated, which include, without limitation, chronic pancreatitis, cryptogenic cirrhosis, epidermolysis bullosa simplex; epidermolytic hyperkeratosis; ichtyosis bullosa of siemens; inflammatory bowel disease; loose anagen hair syndrome; meesman corneal dystrophy; monilethrix; oral white sponge nevus; pachyonychia congenital (pc); palmoplantar keratoderma: including epidermolytic, non-epidermolytic, and striate; pseudofolliculitis barbae; steatocystoma multiplex; and the like, as well as the liver diseases discussed herein.

[56] For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid

preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g., lecithin or acacia); non-aqueous vehicles (e.g., almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate. Preparations for oral administration may be suitably formulated to give controlled release of the active compound. For buccal administration the compositions may take the form of tablets or lozenges formulated in conventional manner.

[57] The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

[58] In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

[59] The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

[60] The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one specific K8/K18 nucleic acid reagent or restriction enzyme described herein, which may be conveniently used, e.g., in clinical settings, for prognosis of patients susceptible to liver disease.

[61] Before the subject invention is described further, it is to be understood that the invention is not limited to the particular embodiments of the invention described below, as variations of the particular embodiments may be made and still fall within the scope of the appended claims. It is also to be understood that the terminology employed is for the purpose

of describing particular embodiments, and is not intended to be limiting. Instead, the scope of the present invention will be established by the appended claims.

[62] In this specification and the appended claims, the singular forms "a," "an" and "the" include plural reference unless the context clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs.

[63] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range, and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[64] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs. Although any methods, devices and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods, devices and materials are now described.

[65] All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the subject components of the invention that are described in the publications, which components might be used in connection with the presently described invention.

EXPERIMENTAL

Methods

[66] *Patients:* We included for the analysis specimens of 467 explanted livers that were obtained from the liver transplantation units at Stanford University, the University of California San Francisco, and California Pacific Medical Center. The 467 liver samples were examined for K8/K18 mutations in 5 exonic regions, that included the epidermal keratin domains where most of the mutations have been identified (Irvine & McLean, Br. J. Dermatol. 140:815-828, 1999). The frequency of K8/K18 mutations were then determined in the remaining 10 exons of K8/K18. Peripheral blood samples from 349 healthy volunteers were obtained from the Stanford Blood Bank and used for genomic DNA isolation. In addition and when available, blood samples were obtained from patients with the identified keratin mutations and/or from their children or parents. The patients' sex and racial/ethnic background were determined from patients' medical records. No information could be found on 15 patients due to lack of

records or to retransplant. For the control samples, anonymous information regarding sex and race was provided by the Stanford Blood Bank. The diagnoses were based on the United Network for Organ Sharing transplant listing. Medical records of all patients with a keratin mutation were reviewed and the diagnosis was confirmed.

[67] *Histopathology and statistical analysis:* Pathology slides from the explanted livers with keratin mutations, and matched liver disease controls, were reviewed by a single pathologist who did not know which specimens harbored the keratin mutations. Slides from the explanted livers of all 17 patients with keratin mutations, as well as disease-matched controls, were examined and scored for the presence or absence of features including Mallory and acidophil bodies, cell size, ground glass cytoplasm, and dysplasia. Images of the hematoxylin and eosin stained liver sections were obtained using a Nikon Eclipse E1000 microscope with a 40x objective. Data analysis was conducted using the Fisher's exact test performed with Statistical Analyzing System software (SAS).

[68] *Molecular methods:* Genomic DNA was prepared using a Dneasy tissue kit (QIAGEN Inc., Chatsworth, CA). Exonic regions (Fig. 1A) were amplified by the polymerase chain reaction. The amplified products were analyzed for mutation using the WAVE® System (Transgenomic Inc, San Jose, CA), and any samples with a "shift" pattern suggestive of a mutation were sequenced in the forward and reverse directions to confirm the presence of a mutation (Fig. 1B).

[69] *Biochemical methods*: Tissues were homogenized in phosphate buffered saline containing 1% n-dodecyl-*N*,*N*-dimethylglycine (Empigen BB, Calbiochem-Novabiochem, San Diego, CA), 5 mM EDTA, and protease inhibitors. Homogenized samples were solubilized for 30 minutes, pelleted then used for immunoprecipitation of K8/K18. Precipitates were analyzed by: (i) SDS polyacrylamide gel electrophoresis (PAGE), under reducing or non-reducing conditions, then Coomassie staining, (ii) SDS-PAGE then immunoblotting, or (iii) two-dimensional gels using isoelectric focusing (horizontal direction) and SDS-PAGE (vertical direction) then immunoblotting.

[70] *Mass spectrometry analysis:* Separated K8 and K18 bands were cut out from preparative gels, reduced with dithiothreitol, alkylated with iodoacetamide, and then digested with trypsin in 50 mM ammonium bicarbonate (pH 7.8) using a standard in-gel-digestion procedure. Extracted K8 or K18 tryptic peptides were desalted using a C18 ZipTip (Millipore, MA) then eluted with 50% acetonitrile-0.1% trifluoroacetic acid (TFA). A 1 µl aliquot of the eluant was mixed with equal volume of matrix solution (saturated a-cyano-4-hydroxycinnamic acid in 0.1% TFA-50% acetonitrile in water) and analyzed by a MALDI-TOF mass spectrometer (Bruker Biflex III) equipped with a nitrogen 337 nm laser. The mass spectra were acquired in the reflectron mode. Internal mass calibration was performed with two trypsin

autodigested fragments (842.5 and 2211.1 Da). K8 tryptic peptides were also digested with CNBr and similarly analyzed.

[71] *Antibody generation:* Two polyclonal antibodies were generated after immunizing rabbits with synthetic peptides containing wild type K8 R340 (^{335}ADAEQ***R***GELAI) or mutant K8 H340 (^{335}ADAEQ***H***GELAI). As shown in Fig. 3B, anti-K8 R340 or anti-K8 H340 antibody recognized predominantly K8 WT or the K8 R340H mutant, respectively, that were co-expressed in transfected BHK cells with wild type K18.

[72] *Immunofluorescence staining:* Snap-frozen liver explants were embedded in optimum cutting temperature compound, sectioned then fixed in acetone (-20° C, 10 min). Sections were double stained with antibodies directed to K8/K18 or vimentin (NeoMarkers, Freemont, CA). Flourescence images were obtained using an MRC 1024ES confocal scanner (BioRad, Hercules, CA) coupled to a Nikon Eclipse TE300 microscope.

Results

[73] *Identification of K8 and K18 mutations:* We tested DNA extracted from liver explants or peripheral blood for the presence of K8 or K18 mutations. Two cohorts were examined, whose demographics are summarized in Table 1: a group of 467 patients with a variety of liver diseases, and a control group of 349 blood bank donors. The ethnic background of the two cohorts was generally similar except for a higher preponderance of Caucasian patients in the control group.

Table 1. Racial/Ethnic Background and Sex of Patients and Controls

Charactertics	Liver Disease Patients		Blood Bank Controls	
White	**274**	**58.6%**	**268**	**76.8%**
Male	139		154	
Female	135		114	
Black	**26**	**5.6%**	**9**	**2.6%**
Male	9		5	
Female	17		4	
Hispanic	**67**	**14.3%**	**42**	**12.0%**
Male	28		21	
Female	39		21	
Asian/Pacific Islander	**57**	**12.1%**	**23**	**6.6%**
Male	36		12	
Female	21		11	
Middle Eastern/Indian	**20**	**4.2%**	**6**	**1.7%**
Male	14		2	
Female	6		4	
Native American	**2**	**0.8%**	**1**	**0.30%**
Male	1		0	
Female	1		1	
Unknown	**21**	**4.4%**	**0**	**0%**
Total	**467**		**349**	

[74] The etiology of the liver diseases is broad (Table 2), most of which is noncryptogenic (based on clinical criteria, 68 of the 467 patients were classified as having cryptogenic liver disease; Table 2). We included the control blood bank cohort in order to address which mutations identified in the liver disease cohort are likely to represent "true" mutations versus polymorphism variants found in the general population. Seven heterozyous missense mutations in 5 exonic regions of K8/K18 were identified in 17 of 467 liver disease patients and 2 of 349 blood bank controls (3.6% and 0.6% frequency; $p<0.004$). The 5 exonic regions included the epidermal keratin domains where most of the mutations have been identified in skin disease patients.

Table 2. Liver Disease Etiologies and Frequency of Keratin Mutations

Etiology of Liver Disease	# of Patients (# with keratin mutations)	% Keratin Mutations
Hepatitis C	80 (9)	11.2
Hepatitis C/Alcohol	28 (5)	17.8
Hepatitis B	44 (5)	11.4
Biliary Atresia	47 (3)	6.4
Alcohol	33 (3)	9.1
Cryptogenic	68 (7)	10.3
Primary Biliary Cirrhosis	12 (4)†	33.3
Primary Sclerosing Cholangitis	15 (1)	6.7
Acute Fulminant Hepatitis	35 (4)†	11.4
Neonatal Hepatitis	9 (2)	22.2
Autoimmune Hepatitis	33 (2)	6.1
Metabolic/Genetic*	23	8.7
Cystic Fibrosis	(1)	
Metabolic/other	(1)	
Primary Liver Cancer**	5 (0)	
Drug induced Liver Failure	9 (1)	11.1
Other§	26 (10)†	38.5
Total	467 (58)	12.4

* Metabolic/genetic diagnoses included 4 Wilson's disease, 3 hemochromatosis, 3 α1-antitrypsin deficiency, 2 cystic fibrosis, 3 primary oxalosis, 1 Crigler-Najjar, 1 ornithine transcarbamoylase deficiency, 1 Nieman-Pick, 1 arginosuccinicaciduria, 1 tyrosinemia, 1 cirtrolemia, 1 glycogen storage, and 1 "metabolic disease" that is unclassified.

** Patients with viral hepatitis and hepatoma were included under the appropriate viral hepatitis category. Primary liver cancers included 1 hepatoma, 3 hemangio-endotheliomas, and 1 hepatosarcoma.

† One patient with primary biliary cirrhosis had a double mutant (K8 R340H and K18 R260Q), one patient with acute fulminant hepatitis had a double mutant (K8 R340H and K18 T102A), and one patient with unknown liver disease etiology had a double mutant (K18 I149V and K18 T294M). All other patient had a single mutation.

§ The "Other" category of liver diseases included Budd-Chiari, hepatic artery thrombosis, polycystic disease, Caroli's disease, Byler's disease, 2 chronic rejection, primary graft nonfunction, Klatskin tumor, parenteral nutrition-induced, carcinoid, hepatitis B and C, hepatitis B and alcohol, multiple adenomas, secondary biliary cirrhosis, veno-occlusive disease, and congenital hepatic fibrosis. At least 11 of the patients screened had received a prior orthotopic liver transplant. For a few patients, their disease etiology could not be discerned from the medical records.

Table 3. K8/K18 Mutations in Human Liver Diseases

	Mutations		# of mutation carriers from	
	Amino Acid SEQ ID NO:4	Nucleotide SEQ ID NO:3	Liver disease cohort (467)	Blood bank controls (349)
K8	**G52V**	**GGC→GTC**	**1**	**None**
	Y53H	**TAT→CAT**	**5**	**1**
	G61C	**GGC→TGC**	**6**	**1**
	R340H	**CGT→CAT**	**30**	**10**
	G433S	**GGC→AGC**	**4**	**1**
	R453C	**CGC→TGC**	**1**	**None**
	1-465(I) RDT(468)	**Frameshift**	**1**	**None**
	I62V*	ATC→GTC	1	7
	L71L**	CTG→CTA	1	None
	A318S*	GCT→TCT	5	2
	R301C*	CGC→TGC	None	1
	E376E**	GAG→GAA	2	-
	V460M*	GTG→ATG	None	1
	V479I*	GTC→ATC	None	2
	Amino Acid SEQ ID NO:2	Nucleotide SEQ ID NO:1	Liver disease cohort (467)	Blood bank controls (349)
K18	**Δ 64-71 (TGIAGGLA)**	**Deletion**	**2**	**None**
	T102A	**ACC→GCC**	**1**	**None**
	H127L	**CAT→CTT**	**2**	**None**
	I149V	**ATC→GTC**	**2**	**None**
	R260Q	**CGG→CAG**	**1**	**None**
	E275G	**GAG→GGG**	**1**	**None**
	Q284R	**CAG→CGG**	**1**	**None**
	T294M	**ACG→ATG**	**1**	**None**
	T296I	**ACA→ATA**	**1**	**None**
	G339R	**GGG→AGG**	**1**	**None**
	S229T*	AGC→ACC	2	4
	Y330Y**	TAC→TAT	1	None
K8			48/467	13/349
K18			13/467	None
K8/18			**58†/467**	**13/349**

Single letter standard abbreviations are used to represent amino acids and nucleotides. Sequences with bold lettering refer to mutations that pose a potential risk factor for subsequent development of liver disease, based on analysis of the liver disease cohort and the blood bank control group. Single asterisks (*) highlight amino acid substitutions that are considered polymorphisms since they were found at similar, or higher, incidence in the control cohort as compared with the liver disease group. Double asterisks (**) highlight "silent" nucleotide mutations that do not result in any amino acid change. The total number of K8/K18 mutations shown represents true mutations and does not include "silent" mutations or polymorphisms. †Three patients had a double mutant (K8 R340H and K18 R260Q, K8 R340H and K18 T102A or K18 I149V and K18 T294M), which results in a total number of patients with keratin mutation of 58 (i.e. 61-3=58).

[75] The 10 remaining exonic regions were also analyzed for K8/K18 mutations. Aside from several new polymorphisms, we identified 10 novel K8/K18 amino acid heterozygous mutations (four K8 and six for K18) caused by one deletion, one frameshift and eight

missense alterations. The K8 frameshift mutation at Ile-465 generated a truncated 468 (instead of 482) amino acid protein. The K18 N-terminal domain deletion (Δ64-71) generated a 421 (instead of 429) amino acid protein. The new K8/K18 mutations were found in 44 of 467 patients and 11 of 349 controls (included in Table 3 as a full listing of all K8/K18 mutations identified to date). Three patients had a double mutant. One patient with primary biliary cirrhosis had K8 R340H and K18 R260Q, one patient with acute fulminant hepatitis had K8 R340H and K18 T102A, and one patient with unknown liver disease etiology had K18 I149V and K18 T294M. These data indicate that the most common mutation is K8 R340H (30 of 58, ~52%; Table 3).

Table 4. Diseases and Ethnicities Associated with Different Keratin Mutations in Liver Patients and Controls

	Liver Disease Patients		Blood Bank Controls
Keratin Mutation	**Liver Disease**	**No. with mutation/Ethnicity***	**No. with mutation/Ethnicity**
K8 G52V	Viral Hepatitis	1/White	none/-
K8 Y53H	Viral Hepatitis, BA, CC	5/3 Black, 1 White, 1 Hispanic	1/Black
K8 G61C	Viral Hepatitis,Alcohol,CC,CF	6/1 Black, 4 White, 1 Unknown	1/White
K8 R340H	Viral Hepatitis, AFH, others	30/14 White, 10 Hispanic, 1 Indian, 5 Unknown	10/8 White, 2 Hispanic
K8 G433S	Viral Hepatitis, Alcohol, BA	4/2 Black, 2 Unknown	1/Black
K8 R453C	CC	1/Unknown	none/-
K8 1-465(I)RDT(468)	PSC	1/White	none/-
K18 Δ 64-71 (TGIAGGLA)	Viral Hepatitis/Alcohol	2/1 White, 1 Unknown	none/-
K18 T102A	AFH	1/Hispanic	none/-
K18 H127L	Metabolic, CC	2/White	none/-
K18 I149V	PBC	2/1 Hispanic, 1 Unknown	none/-
K18 R260Q	PBC	1/Hispanic	none/-
K18 E275G	Viral Hepatitis	1/White	none/-
K18 Q284R	Viral Hepatitis	1/White	none/-
K18 T294M	Unknown	1/Unknown	none/-
K18 T296I	Viral Hepatitis	1/Asian	none/-
K18 G339R	Hepatic Artery Thrombosis	1/Unknown	none/-
Prevalence of Keratin Mutations		58**/467 (12.4%)†	13/349 (3.7%)†

BA= Biliary Atresia, CC= Cryptogenic Cirrhosis, CF= Cystic Fibrosis, PBC= Primary Biliary Cirrhosis, AFH= Acute Fulminant Hepatitis, PSC=Primary Sclerosing Cholangitis

* The ethnicities and sexes of the 58 liver disease patients with keratin mutations were 26 White (11 female/15 male), 6 Black (4 female/2 male), 12 Hispanic (8 female/4 male), 1 Indian (1 female), 1 Asian (1 male) and 12 unknown.

** Three patients had a double mutation (i.e. these 58 patients are counted only once, see also legend for Table 2). One patient with primary biliary cirrhosis had K8 R340H and K18 R260Q, one patient with acute fulminant hepatitis had K8 R340H and K18 T102A, and one patient with unknown liver disease etiology had K18 I149V and K18 T294M.

†P value<0.0001 (Proportion of patients with keratin mutations from the combined liver disease cohorts compared with blood bank controls).

[76] The mutations that were identified in 58 of 467 patients represent a mutation frequency of 12.4%, as compared with a mutation frequency of 3.7% found in 13 of 349 controls (P<0.0001; Tables 3 and 4). Given the demographics of patients with keratin mutations, there does not appear to be any obvious accumulation of keratin mutations in a particular sex or ethnic background (Table 4). Since most of the specimen that we analyzed consisted of liver explants, we also tested blood specimen for the germline presence and transmission of the identified keratin mutations, in order to exclude the possibility that the mutations we identified

occurred during development of disease. Of the 58 independent patients with K8/K18 mutations we were able to locate 21 patients and/or their offspring (12 patients and 9 offsprings). Nineteen of these patients and/or their offspring blood specimen had the identical heterozygous keratin mutation to that identified in the diseased explanted liver, and two of the offsprings had wild type K8 (Table 5). Therefore, the K8/K18 mutations we identified are not a consequence of the liver disease but, rather, predispose their carriers to subsequent development of liver disease.

Table 5. Germline Presence and Transmission of K8/K18 Mutations*

Keratin mutation		Liver disease	Mutation in blood of	
			patients	offspring
K8	Y53H	Hepatitis B	Yes	N.A.
	G61C	Cryptogenic	Yes	Yes** (in 3 of 4 tested)
	R340H	Idiopathic FHF	Yes	N.A.
	R340H	Hepatitis C and alcohol	Yes	N.A.
	R340H	Hepatitis C	Yes	Yes** (in 1 of 2 tested)
	R340H	Biliary atresia	Yes	N.A.
	R340H	Hepatitis B	Yes	N.A.
	R340H	Cryptogenic	Yes	N.A.
	G433S	Biliary atresia	Yes	N.A.
	1-465(I) RDT(468)	PSC	Yes	N.A.
K18	Δ 64-71 (TGIAGGLA)	Hepatitis C and alcohol	N.A.	Yes (1 tested)
	H127L	Cryptogenic	N.A.	Yes (1 tested)
	I149V	PSC	Yes	N.A.
	R260Q	PBC	Yes	Yes (1 tested)

FHF= Fulminant hepatic failure, PBC= Primary biliary cirrhosis, PSC=Primary sclerosing cholangitis

N.A.= not available

*All identified mutations are heterozygous. ** One of the tested offspring has wild type K8.

[77] *Expression of mutant keratin proteins:* We examined, biochemically, the liver explant specimens from some of the mutation carriers within the cohort of 467 patients in order to confirm the presence of the mutation at the protein level. Mutations that were examined included 8 types: K8 G52V/Y53H/G61C/R340H and K18 T102A/H127L/R260Q/G339R (Table 3). The presence of the K8 G61C protein was confirmed as we did previously (Ku et al, N Eng J Med 344:1580-1587, 2001), by formation of K8 dimers (under nonreducing gel conditions) due to the newly introduced cysteine (normally absent in K8/K18). We also used two-dimensional gels to separate proteins based on their charge, and thereby confirmed the presence of the K8 Y53H and K18 H127L/R260Q/G339R species (Fig. 2A). These four K8/K18 mutations generated proteins with a different charge as compared to their wild-type counterparts, and resulted in new isoforms (4 isoforms instead of 2 for mutant K8, and 5 isoforms instead of 3 for mutant K18, Fig. 2A).

[78] Two-dimensional gel analysis was, however, not informative for mutations that do not significantly alter the isoelectric point (Fig. 2A, K8 G52V and K18 T102A variants). For these mutations, we compared the mass spectrometric profiles of protease-generated fragments of wildtype and mutant keratins and tested for the presence of peptides that have a mutation-altered mass. As shown in Fig. 2B, presence of the K18 T102A protein was confirmed by detection of its alanine-102-containing peptide with a predicted mass of 818.3 (in addition to the wild-type threonine-102-containing peptide with a predicted mass of 848.3). A similar analysis was attempted for K8 G52V but no peptide that corresponds to a glycine-to-valine substitution was detected, likely due to inability to recover it from the isolation column.

[79] Given that the K8 R340H mutation was the most common mutation, found in 30 of 58 affected individuals at the highly conserved K8 R340 (Fig. 3A), we generated and used polyclonal antibodies to examine the presence of wild type K8 versus the mutant K8 R340H protein in patient liver explants. The two antibodies we generated were directed towards the wild type K8 R340 residue (SEQ ID NO:5) (^{335}ADAEQ*R*GELAI) or the mutant K8 H340 residue (SEQ ID NO:6) (^{335}ADAEQ*H*GELAI). Specificity of the antibodies was tested in transfected cells that expressed K8 WT or K8 R340H. Anti-K8 R340 (i.e. K8 WT) or anti-K8 H340 antibody recognized predominantly K8 WT or the K8 R340H mutant, respectively (Fig. 3B). Expression of the K8 R340H protein in patient liver explants was confirmed by immunoblotting of total explant liver lysates with the antibodies. As shown in Fig. 3C, anti-K8 R340 (WT) recognizes K8 in the control patient and the patients with K8 R340H (lanes 1-6), whereas anti-K8 H340 (mutant) recognizes only patient livers with the K8 R340H mutation (lanes 2-6) but not the normal liver (lane 1). This indicates that the patients with K8 R340H mutation are heterozygous.

[80] *Effect of keratin mutations on keratin filament organization and liver histology:* We compared keratin filament organization in liver explants of patients with and without keratin mutations. We did not observe any generalized keratin mutation-specific organization defects as tested by immunofluorescence staining (Fig. 4A). The diseased livers (with or without keratin mutations) had reorganization of the keratins filaments with the most prominent feature being thickening and partial collapse (Fig. 4A, panels d and g) as compared with normal liver keratin staining (Fig. 4A, panel a). The diseased livers had variable but significant vimentin-positive staining (Fig. 4A, panels e, h), which was used as a fibroblast/stellate cell marker, as compared with normal liver (Fig. 4A, panel b). Vimentin staining did not correlate with the presence of keratin mutations, and did not involve hepatocytes (Fig. 4A; merged images c, f, i). Analysis of additional liver samples with proper attention to sample handling will be needed to better assess any potential keratin mutation-induced effects on keratin organization.

[81] We also asked if any histologic features identified by light microscopy could distinguish cirrhotic livers of patients with and without keratin mutations. Features such as Mallory's

hyaline, acidophil bodies, enlarged hepatocytes, ground glass cytoplasm and dysplasia were found in keratin-mutant and non-keratin-mutant livers. However, close inspection of the liver specimen(s) showed a unique accumulation in some hepatocytes of cytoplasmic filamentous arrays (Fig. 4B). When coding of the slides (with mutant or nonmutant keratin) was opened, the findings indicated that the filamentous deposits were found in 10 of 17 tested patients with keratin mutations but in only 3 of 16 disease-matched controls ($p=0.03$). When patients with only primary hepatocellular diseases (cryptogenic, viral hepatitis, alcohol, acute fulminant hepatitis) were included in the analysis, 10 of 11 explants with a keratin mutation contained the cytoplasmic filamentous deposits as compared with 3 of 13 disease-matched controls ($p=0.001$). The nature of the filamentous deposits remains to be determined, but they do not appear to correspond to aggregated keratins since they were not recognized by anti-keratin antibodies (which may reflect epitope masking).

Discussion

[82] Mutations in keratins and other IF family members, including lamins, desmin, glial fibrillary acidic protein (GFAP) and neurofilaments, are well-established causes of a wide range of tissue specific human diseases. The list of newly identified diseases associated with IF proteins continues to grow, including the latest association of K8 with liver disease, GFAP with Alexander disease and the neurofilament-L chain with Charcot-Marie-Tooth type-2. One distinguishing feature of K8/K18 mutations, as compared with epidermal keratin mutations involving K5/K14/K1/K10, is that the epidermal keratin diseases are typically autosomal dominant with ~100% penetrance. In contrast, K8/K18 mutations appear to be risk factors with variable penetrance, rather than direct causes of disease. In support of this, the location of the characterized K8/K18 mutations does not involve conserved pan-keratin domain mutation hot spots that have been identified in epidermal keratins (Fig. 5). Apparent absence of such mutations suggests that they may be lethal, given that K8/K18 are among the earliest expressed keratins during embryogenesis. Also, K8/K18 mutations are germline and do not simply result as a consequence of the liver disease (Table 5).

[83] The ages of the offspring with the keratin mutations range from 31-52, but none of these carriers have apparent liver disease, based on clinical history and serologic testing. These observations support a "multi-hit" hypothesis, whereby one major "hit" is carrying a relevant K8/K18 mutation with subsequent "hits" including underlying liver disease or exposure to injurious factors such as toxins or viruses (Fig. 6). Clinical and natural history studies can be used to define the relative risk of subsequent development of liver disease, or the relative increase in progression of an underlying liver disease, in those who carry specific K8/K18 mutations.

[84] At the molecular level mutations in IF chains can, in principle, alter the α-helical propensity of the chains, the number and type of the intra- and inter-helical ionic interactions,

increase or decrease the stability of the α-helical strands, modify the helix capping potential and change the hydrogen bonding ability of the chains. Mutations could also adversely affect the ability of the molecules to assemble in to viable IF, to bundle as normally required, or to function properly even if assembled correctly (Table 6). Therefore, potential effects of the keratin mutations include destabilizing K8/K18 filaments, interrupting ionic interactions, introducing disulfide bonds, or altering keratin phosphorylation/solubility. These molecular consequences of keratin mutations may interfere the normal filament reorganizations that occur in hepatocytes upon multiple physiologic and nonphysiologic stimuli, and ultimately result in liver disease (Fig. 6).

Table 6. Molecular Consequences of Keratin mutations

Mutations		Potential effects
K8	R340H	Destabilization
	G433S	Altering keratin phosphorylation
	R453C	Formation of a disulfide bond
	1-465(I) RDT(468)	Destabilization
K18	Δ 64-71(TGIAGGLA)	Destabilization
	E275G	Destabilization
	Q284R	Destabilization
	T294M	Interruption of ionic interaction
	T296I	Interruption of ionic interaction

[85] The significant number of patients described in this study with K8/K18 mutations provide several insights into keratin-associated liver diseases. For example, K8 Y53H, K8 G61C, and most prominently K8 R340H are shown to be mutation hot spots. These mutations were found in 5, 6 and 30 of the 58 patients with K8/K18 mutations, respectively (Table 3 and Fig. 5), and collectively make up ~71% of the K8/K18 mutations identified to date, while the K8 R340H mutation alone makes up ~52% of all K8/K18 mutations identified to date. Analysis of additional liver disease patients will help determine if these mutation hot spots maintain their frequency. Also, search for additional keratin mutation carriers in a broad range of cryptogenic and noncryptogenic liver diseases is clearly warranted.

[86] When we initially identified K8 and K18 mutations in 6 patients, all 6 had cryptogenic liver disease (Ku et al, J Clin Invest 99:19-23, 1997 and Ku et al, N Engl J Med 344:1580-1587, 2001). This liver disease makes up nearly 10% of all patients who undergo liver transplantation. In this cryptogenic liver disease cohort 7 of 68 patients (10.3%) have K8 or K18 mutations. However and more importantly, our present and more significant findings are that keratin mutations are also common (12.8%) in the noncryptogenic liver disease group (Table 2) that accounts for the remaining 90% of liver transplantations (based on the patient groups we studied). The association of keratin mutations with cryptogenic liver disease raises the possibility that some diseases that are linked with this type of cirrhosis, such as

nonalcoholic steatohepatitis, may also be associated with keratin mutations. More K8 and/or K18 mutations may still be identified.

[87] Although the mechanisms by which keratin mutations predispose to cirrhosis remain to be defined, already known and emerging keratin functions are likely to be involved. For example, multiple transgenic mouse model studies showed that K8/K18 serve the essential function of protecting hepatocytes from a variety of stresses including agents that cause acute (e.g. acetaminophen) or chronic (e.g. griseofulvin) injury, and agents that induce apoptosis (e.g. Fas antibody). K8/K18 may also be involved in protein targeting to the apical compartment of polarized epithelia, interacting with apoptotic machinery proteins, cell signaling and regulating the availability of abundant cellular proteins. Hence, keratin mutations may potentially act at a number of functional cellular nodes. One surrogate marker of keratin function is cytoplasmic filament organization, which was shown to be abnormally altered, only after stress exposure, in the K8 Y53H/G61C mutations (Ku et al, N Engl J Med 344:1580-1587, 2001). Our observation of preferential cytoplasmic filamentous deposits in cirrhotic livers of patients with keratin mutations is likely to be relevant and is reminiscent of Rosenthal fibers that are seen in association with Alexander Disease. The nature and pathogenesis of these deposits and their association with keratin-related liver disease remain to be investigated, but they are morphologically distinct from Mallory body-type deposits.

[88] It is evident that subject invention provides a convenient and effective way of determining whether a patient will be susceptible to liver disease. The subject methods will provide a number of benefits, including preventive treatment and diet. As such, the subject invention represents a significant contribution to the art.